Battle of Neighborhoods – Report

Introduction & Problem Statement

When a young family moves to a new area, a number of factors go into choosing the right neighborhood/town for the family to move to.

This project assesses the areas in San Francisco city where the family is moving to and tries to identify the neighborhood that meets the family priorities.

The neighborhood criteria evaluated include -

- Housing cost and availability
- Crime rate
- Schools in the vicinity
- Venues such as parks, restaurants

Data Sources

1. Crime data – The crime data is taken from police incidents reports of 2018 and 2019 from sfgov page. https://data.sfgov.org/Public-Safety/Police-Department-Incident-Reports-2018-to-Present/wg3w-h783

Crime data includes incidents by category and by neighborhood. The data grouped by neighborhood for analysis. The results are plotted using Folium choropleth feature and the neighborhoods with lowest crime rates are chosen.

2. Housing data – The current houses for sale are obtained from Redfin site. http://www.redfin.com.

The data included for the family interest is limited to 2 bedroom houses.

3. Public school listing – The list of public schools with their locations are also available at sfgov page. https://data.sfgov.org/Economy-and-Community/Schools/tpp3-epx2 .

Only elementary schools are included in the analysis.

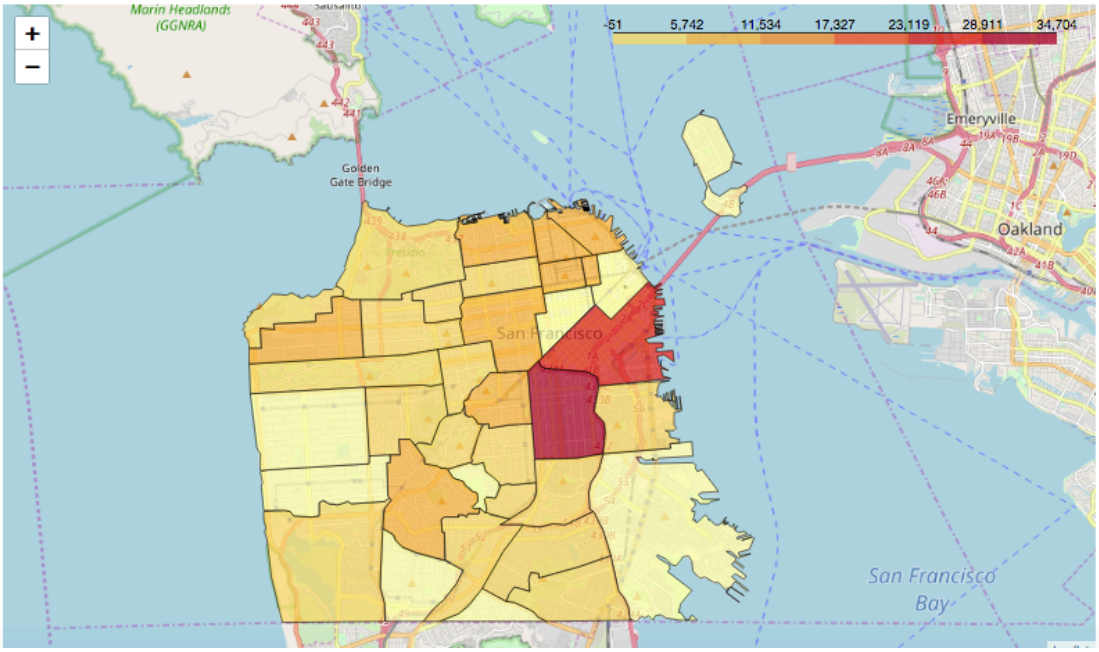
4. Venue details and distribution across neighborhoods is sourced from Foursquare API.

A combination of all the above data is used to narrow and select neighborhood that meets the family's priorities.

Analysis Methodology and Results

Starting with the crime data for the city – the crime incidents are grouped by neighborhood and sorted by total number of incidents. A choropleth of the crime map highlights the high crime areas. High crime neighborhoods are excluded from the analysis going forward and top 20 safe neighborhoods are included.

Analysis Neighborhood	Count	Latitude	Longitude
Mission	34363	37.761487	-122.416917
Tenderloin	30893	37.783283	-122.414454
Financial District/South Beach	28104	37.789309	-122.401061
South of Market	25972	37.778262	-122.407260
Bayview Hunters Point	17097	37.733019	-122.390965
Western Addition	9771	37.782292	-122.428450
North Beach	9675	37.805058	-122.411099
Castro/Upper Market	9015	37.763161	-122.432725
Sunset/Parkside	8680	37.749499	-122.490828
Nob Hill	8535	37.789999	-122.416184

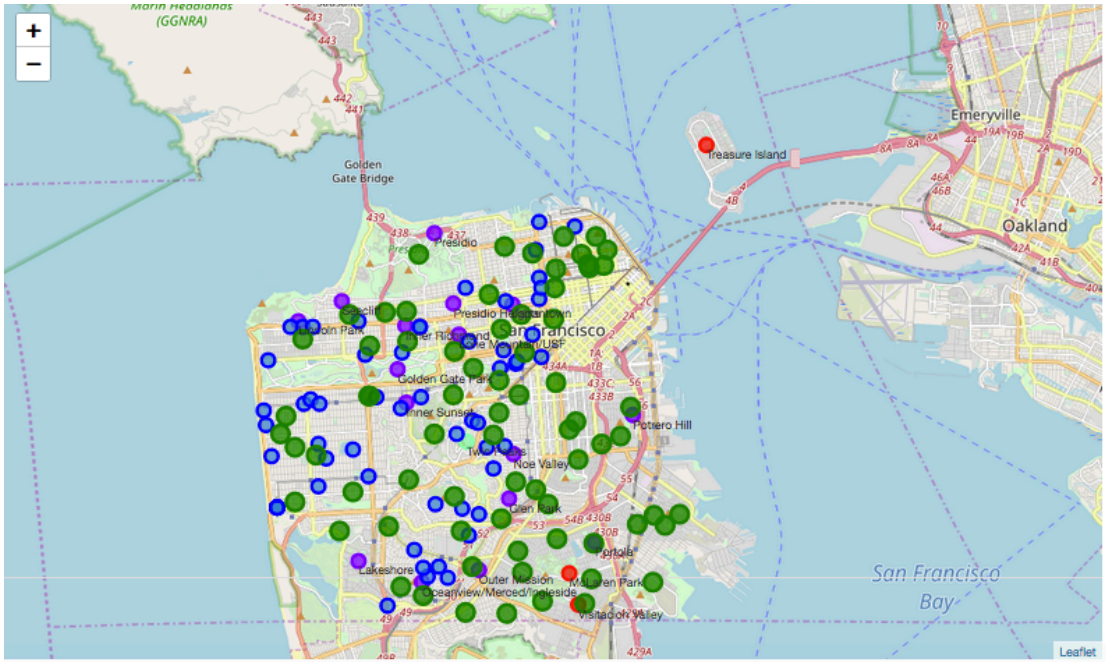


Secondly, FourSquare API is used to gather venue details on the filtered neighborhood set. The venues are grouped by neighborhood after which the frequency of each venue category is calculated. Each neighborhood with its top 10 venue categories are saved to venue dataframe.

	Analysis Neighborhood	Count	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue
17	McLaren Park	290	37.719865	-122.414827	0	Dog Run	Park	Baseball Field	Trail	Yoga Studio	Field	French Restaurant	Fountain
32	Seacliff	330	37.786285	-122.484768	1	Grocery Store	Restaurant	Café	Chinese Restaurant	Coffee Shop	Pizza Place	Sushi Restaurant	Pet Store
14	Lincoln Park	370	37.781395	-122.498410	1	Outdoor Sculpture	Golf Course	Playground	Burmese Restaurant	Martial Arts Dojo	Sandwich Place	Café	Moroccan Restaurant
29	Presidio	730	37.802747	-122.456258	1	Museum	Food Truck	Café	Gym	Sporting Goods Shop	Street Food Gathering	Field	Gift Shop
36	Treasure Island	956	37.824153	-122.372701	0	Park	Island	Bus Station	Flea Market	Grocery Store	Gym	Athletics & Sports	Breakfast Spot
6	Glen Park	1496	37.738044	-122.433462	1	Coffee Shop	Trail	Bakery	Café	French Restaurant	Bubble Tea Shop	Breakfast Spot	Park
37	Twin Peaks	1511	37.751949	-122.446489	2	Scenic Lookout	Trail	Mountain	Yoga Studio	Fast Food Restaurant	Food Truck	Food & Drink Shop	Food
30	Presidio Heights	1714	37.785568	-122.450474	1	Coffee Shop	Café	New American Restaurant	Spa	Cosmetics Shop	American Restaurant	Italian Restaurant	Bed & Breakfast
21	Noe Valley	2725	37.749092	-122.432114	1	Coffee Shop	Bookstore	Bakery	Gift Shop	Gym	Italian Restaurant	Breakfast Spot	Pizza Place
23	Oceanview/Merced /Ingleside	3059	37.717666	-122.460268	1	Playground	Japanese Restaurant	Bus Station	Liquor Store	Chinese Restaurant	Dive Bar	Dog Run	Fountain
7	Golden Gate Park	3257	37.769738	-122.467641	1	Garden	Science Museum	Exhibit	Park	Gift Shop	Food Truck	Lake	Fountain
12	Japantown	3259	37.785463	-122.432440	1	Gift Shop	Bakery	Japanese Restaurant	Tea Room	Café	Grocery Store	Ice Cream Shop	Shopping Mall

Clustering is utilized to group neighborhoods by similar categories. Though this does not yield any clear grouping of the neighborhoods, the dataframe with the venue information and the crime data combined gives a means to narrow down the preferred neighborhoods to –SeaCliff, Lincoln Park,Glen Park, Oceanview, Portola.

Plotting the schools(green circles) and housing data(blue circles) along with the venue clusters enables further filtering of neighborhood of choice from step 2. Glen Park and Portola does not enough housing availability.



The top neighborhoods of choice for the family to move to are –

Lincoln Park,
Oceanview/Merced/Ingleside and
Seacliff

Discussion

There is scope to further categorize the crimes into violent or otherwise and analyze neighborhoods by specific crime type. The housing and crime data could be extended to include more historic data. The venue interests of the family could also be further defined. And the distance of school from chosen housing could also be included as a specific criterion.

Conclusion

This analysis takes the data available publicly about San Francisco neighborhoods and assesses the suitability of them for a family moving in. The crime information gives a very good first indicator of a neighborhood's favorability for the family. Combining the other factors of housing, schools and venues nearby, the top 3 neighborhoods of choice for the family are Lincoln Park, Oceanview/Merced/Ingleside and Seacliff.